PJIM PARSONS JOURNAL FOR INFORMATION MAPPING

THE PARSONS INSTITUTE FOR INFORMATION MAPPING
68 Fifth Avenue
New York, NY 10011
212 229 6825
piim.newschool.edu

No Country for Fat Men

AARON LAI, M.Sc.
THOMAS HO, B.S.
RYAN WALKER, M.A.
ALLISON SULLIVAN, Ph.D.

KEYWORDS *CHIS, interaction visualization, lifestyle, obesity, race, sugar, Tableau Software*

ABSTRACT Visual analytics is gaining importance due to the explosion of data availability and processing capabilities. This growing field has the potential to inform a wide array of disciplines. In this article, we demonstrate a social science application of the data visualization software Tableau by performing an analysis of obesity in California. Extensive research has documented the increasing levels of obesity in the U.S. and the implications for public health. Most of these studies use some variation of a regression analysis, which, although useful for traditionally statistical relationships, can make detecting new relationships difficult. Visual analytics provides a potential solution to explore how formerly unseen attributes may be associated with each other. We explore the relationship between obesity and health behaviors using data from the California Health Interview Survey (CHIS). CHIS collects extensive information on socio-demographic background and health behaviors. Our analysis focuses on three key health behaviors: vegetable, soda, and French fry consumption. We explore how these variables interact with gender, age, race, and several other characteristics by body mass index (BMI).

We find notable differences in health behaviors by gender, level of obesity, and race. The visual analytics tool helped us to identify various factors related to obesity that merit further investigation. More broadly, our demonstration provides an excellent example of how visual analytic tools can empower end-users to find interesting relationships within a morass of data.

INTRODUCTION

Health researchers are benefitting from the rapid increase in the availability of high quality, publicly available data. Government funded data collection efforts, such as the National Health and Nutrition Examination Survey and the National Health Interview Survey, allow researchers to explore the relationships between a wide range of variables and health outcomes. However, with so much data available, summarizing and conducting exploratory data analysis has become much more challenging. Here, we illustrate how visual analytics can be a valuable tool for uncovering unexpected relationships in many-variable health data sets. We use a readily available commercial product called Tableau to perform a visual analysis of factors potentially related to obesity and diabetes prevalence in California.

According to the Centers for Disease Control and Prevention, in 2009-2010, more than one-third of U.S. adults were obese.[1, 2] Extreme levels of obesity are associated with significantly higher levels of mortality relative to the non-obese.[3] Type II diabetes, which is highly co-morbid with obesity, has also reached an epidemic level [3]. Little is known, however, about the causes and trends in obesity on a population level.[4] The prevalence of obesity is related to many different factors, and, therefore, it can be quite difficult to untangle the relationships between obesity and environmental, socioeconomic, and lifestyle variables. Using data visualization techniques to analyze multiple factors leading to obesity simultaneously, we can readily identify potentially important relationships.

BACKGROUND

PUBLIC HEALTH ISSUES AND CHIS DATA

The California Health Interview Survey (CHIS) is a random-dial telephone survey[5] conducted by the UCLA Center for Health Policy Research in collaboration with the California Department of Public Health and the Department of Health Care Services.[6] The study covers a wide range of topics including demographics, lifestyle factors, health status, access to healthcare, and health insurance status. In this article, we focus on the relationship between obesity and general health status, dietary intake, physical activity, insurance status, and demographics. Given the breadth of information in the CHIS data, this survey is an ideal candidate for visual analytic analysis. Using Tableau to generate quick visualizations of key variables, anyone could replicate our results with minimal software skills.

RELATED WORK ON OBESITY

In the U.S., there are documented differences in the prevalence of obesity by demographic characteristics.[7, 8] More men than women are overweight, but more women than men are obese.[9] The prevalence of obesity increases with adult age until age 80.[10] Among men, there are no differences in the prevalence of obesity by race and ethnicity; however, for women, white women are much less likely to be obese than Mexican-American or non-Hispanic Black women.[11] In a diverse state, such as California, these differences are extremely important.

Previous research has explored variables related to obesity, although a full understanding of the causes is lacking. Some of the propensity for being obese may be genetic.[12] However, behaviors have a strong relationship with weight change. Consuming fruits and vegetables is associated with loss, and consuming French fries and sugary beverages are both strongly associated with weight gain.[13] Patterns of obesity also vary with socioeconomic status. Those with higher levels of education and income are less likely to be to be overweight than those with less education or income.[14]

APPROACH AND DESIGN

We downloaded the 2009 adult CHIS dataset from the askCHIS site and converted the SAS dataset into a Tableau data extract. The CHIS data set includes a series of weights, which are variables used to rebalance survey data to reflect the known demographics of the overall population.[15] The CHIS data are weighted to be representative of the non-institutionalized California population.

Because our purpose was to perform a visual analytics and not a detailed statistical analysis, we approached the problem in a simple way. We grouped some variables (e.g., race) for simplicity. Our target audience is healthcare business people; thus, our focus was to reveal relationships in a self-explanatory way. We added as much information to the charts as possible while maintaining a clean design.

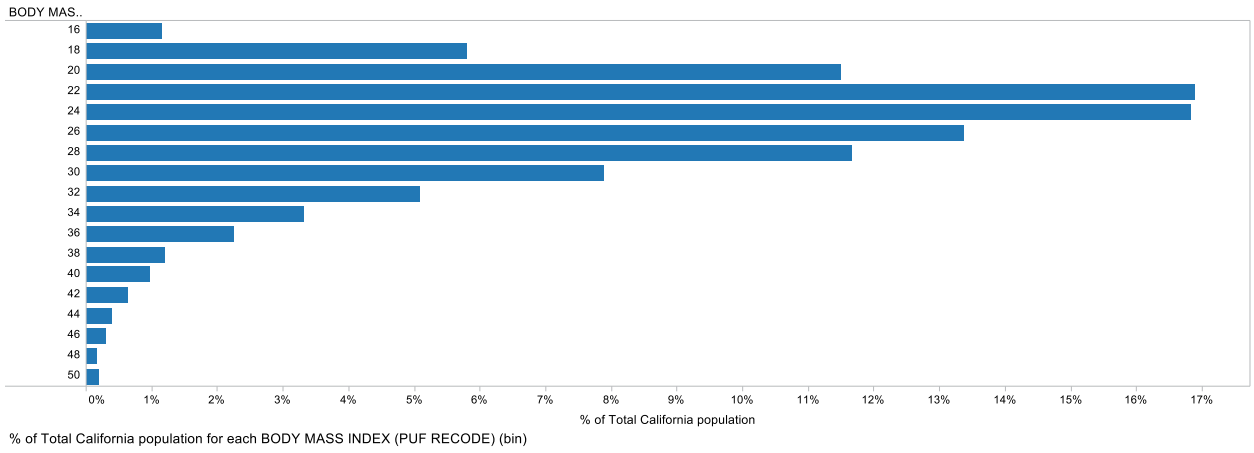
RESULTS

BMI (Body Mass Index) is an individual's weight in kilograms divided by height in meters squared. Although some controversy exists over what BMI is best for health and longevity, the World Health Organization (WHO) describes the optimal BMI to be between 18.5 and 24.9.[9] A BMI of 25–29.9 is considered overweight, and 30 and higher is considered obese. For example, an individual who is 5'7" tall is at healthy weight between 118 and 159 pounds, overweight between 160 and 191 pounds, and obese above 191 pounds.

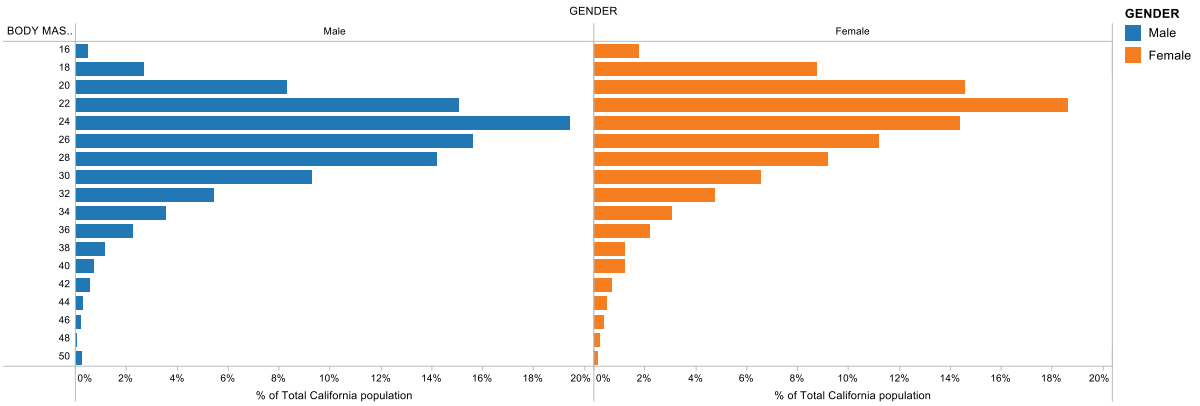
FIGURE 1 is an example of how Tableau can visually delve into multidimensional data; the figure shows the BMI of Californians across various dimensions. PANEL A shows the distribution of BMI among all adults in California. In California, 70.2% of adults have a BMI between 20 and 30. PANEL B shows the distribution of weight by gender. The left side shows male BMI, and the right side shows female BMI. The side by side comparison of these figures shows that females have a lower BMI than males, consistent with previous research. Less consistent with previous research, we do not observe more obese (BMI≥30) females than males. The third panel shows BMI by gender and age. Male BMI is displayed above female BMI. These figures show that females have a markedly lower BMI at lower ages, particularly ages 18–25, but this gender pattern weakens with age.

FIGURE 1 (PANEL A–PANEL C): *Body Mass Index among Californians, California Health Interview Survey, 2009*

PANEL A: *All California adult population*

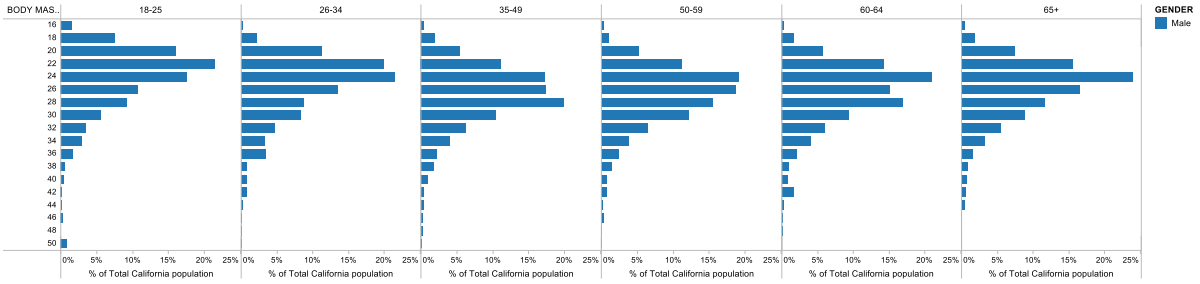


PANEL B: *All California adult population by gender*

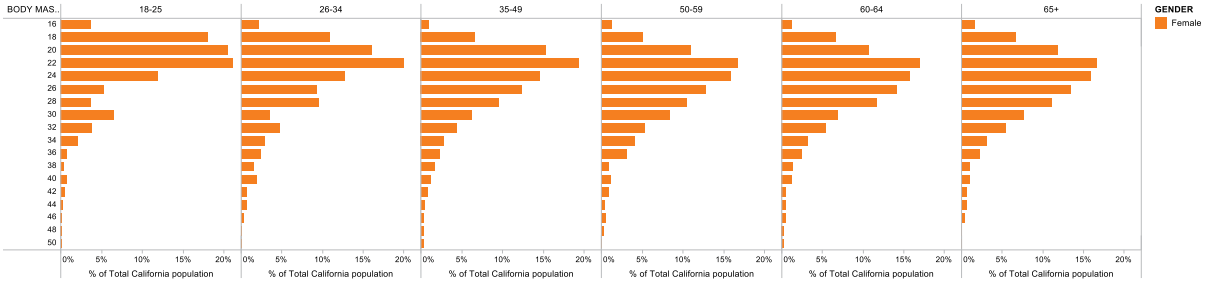


% of Total California population for each BODY MASS INDEX (PUF RECODE) (bin) broken down by GENDER. Color shows details about GENDER

PANEL C: *All California adult population by gender and age*



% of Total California population for each BODY MASS INDEX (PUF RECODE) (bin) broken down by Age. Color shows details about GENDER. The view is filtered on GENDER, which keeps Male



% of Total California population for each BODY MASS INDEX (PUF RECODE) (bin) broken down by Age. Color shows details about GENDER. The view is filtered on GENDER, which keeps Female

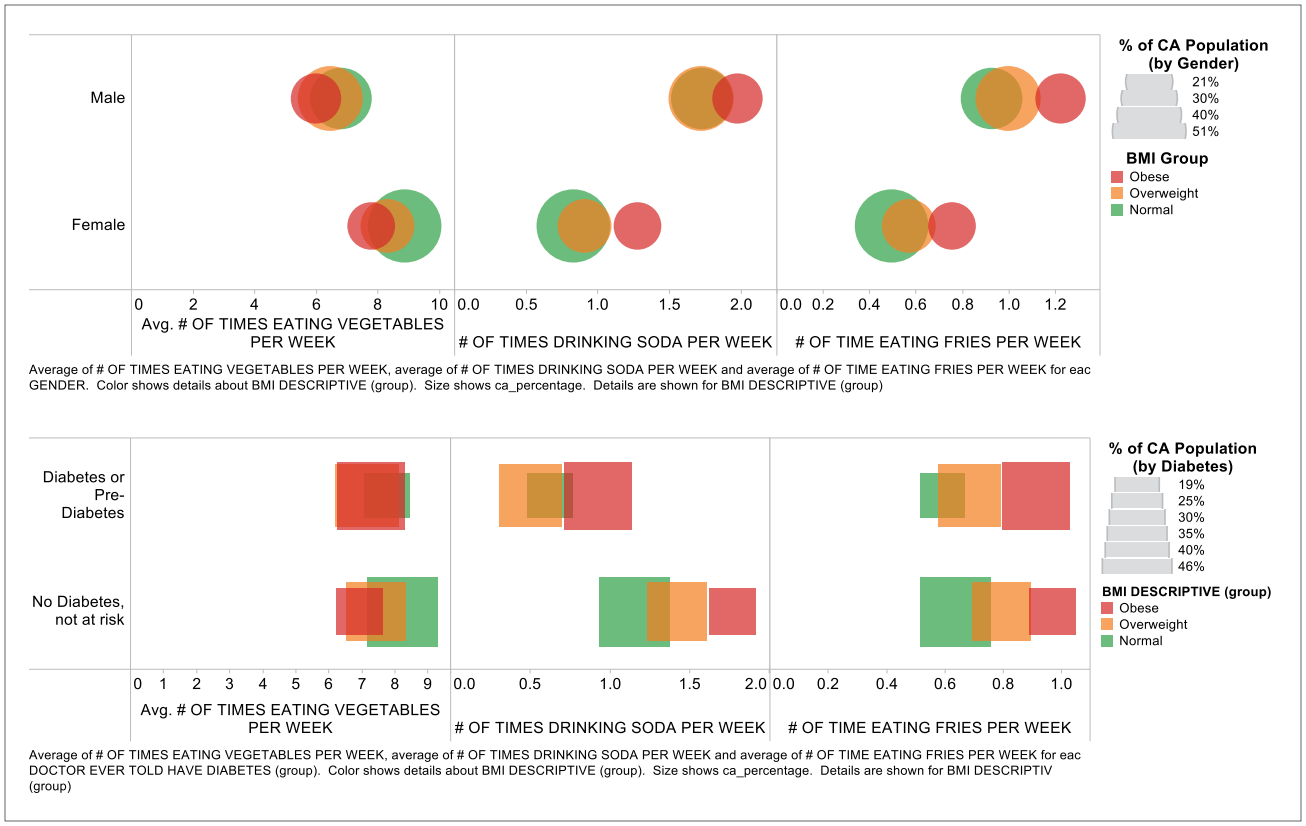


FIGURE 2: *Food preference, gender, and diabetes by weight group, California Health Interview Survey, 2009*

What health behaviors influence BMI by gender? The top panel of FIGURE 2 indicates that men eat fewer vegetables, drink more soda, and eat more French fries than women. The color of the circles denotes BMI group; the size of the circles indicates the proportion of individuals in the particular BMI category by gender. As shown by the red and orange circles, respectively, obese and overweight people of both genders eat fewer vegetables and more French fries than the normal weight group.[16] Overweight and normal women drink nearly the same quantity of soda, but there is a notable shift towards higher soda consumption for both genders in the obese group.

The bottom figure of FIGURE 3 shows that obese people comprise the majority of California diabetics (43%), trailed by the overweight (37%), and normal weight (20%). Diabetics, fortunately, drink significantly less soda than non-diabetics. Nevertheless, obese people drink more soda than normal regardless of diabetes status.

FIGURE 3 shows the distribution of BMI by race and ethnicity. The top row, in red, shows the distribution for African-Americans. As noted in previous research, African-Americans have, on average, a higher BMI than many other groups; nearly half are obese. The distribution of BMI is also high for Hispanics. Alternatively, Asians have the lowest BMI distribution of all groups.

So what health behaviors may influence these race and ethnicity differences in BMI? FIGURE 4 shows BMI by race and ethnicity and health behavior. The size of the diamonds indicates the proportion of individuals of the given race in a given BMI category. The first column shows the average number of times respondents reported eating vegetables per week. Normal weight White respondents reported the highest level of vegetable consumption; obese Latino respondents consumed the fewest vegetables. For all race/ethnicity groups except African Americans, the normal weight group consumed the most vegetables. The next column shows soda consumption. For White Californians, soda consumption and BMI group behave as expected: smaller people consume less soda. For other groups, no consistent pattern between soda consumption and BMI exists. The last column shows fry consumption. For Whites, African Americans, Asians, and other races, more fries are associated with higher levels of BMI.

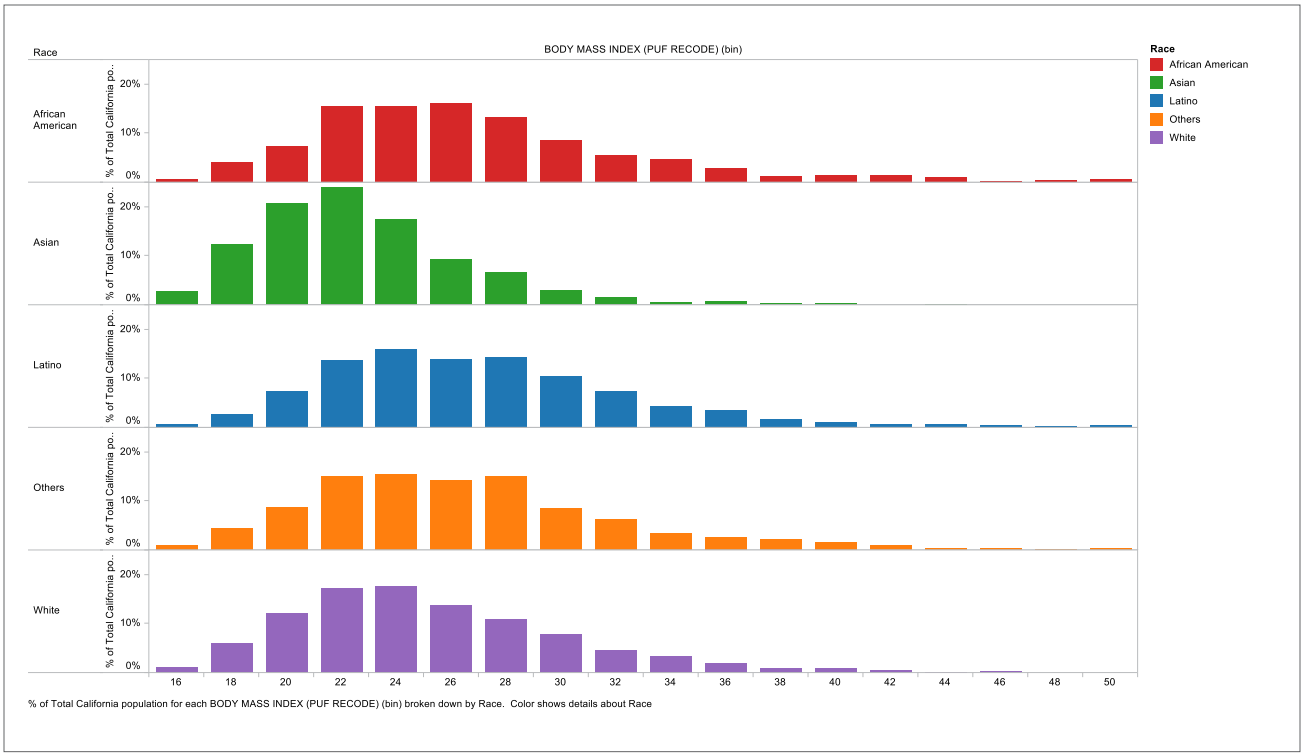


FIGURE 3: *Body Mass Index distribution by race and ethnicity, California Health Interview Survey, 2009*

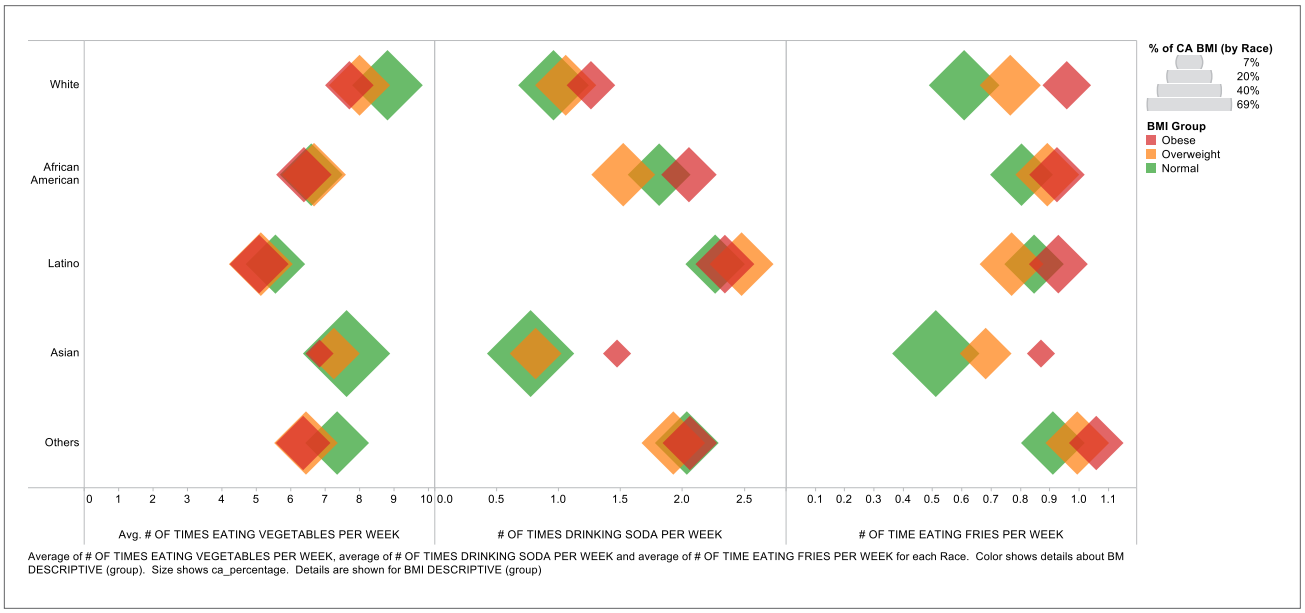


FIGURE 4: *Obesity and food consumption by race and ethnicity, California Health Interview Survey, 2009*

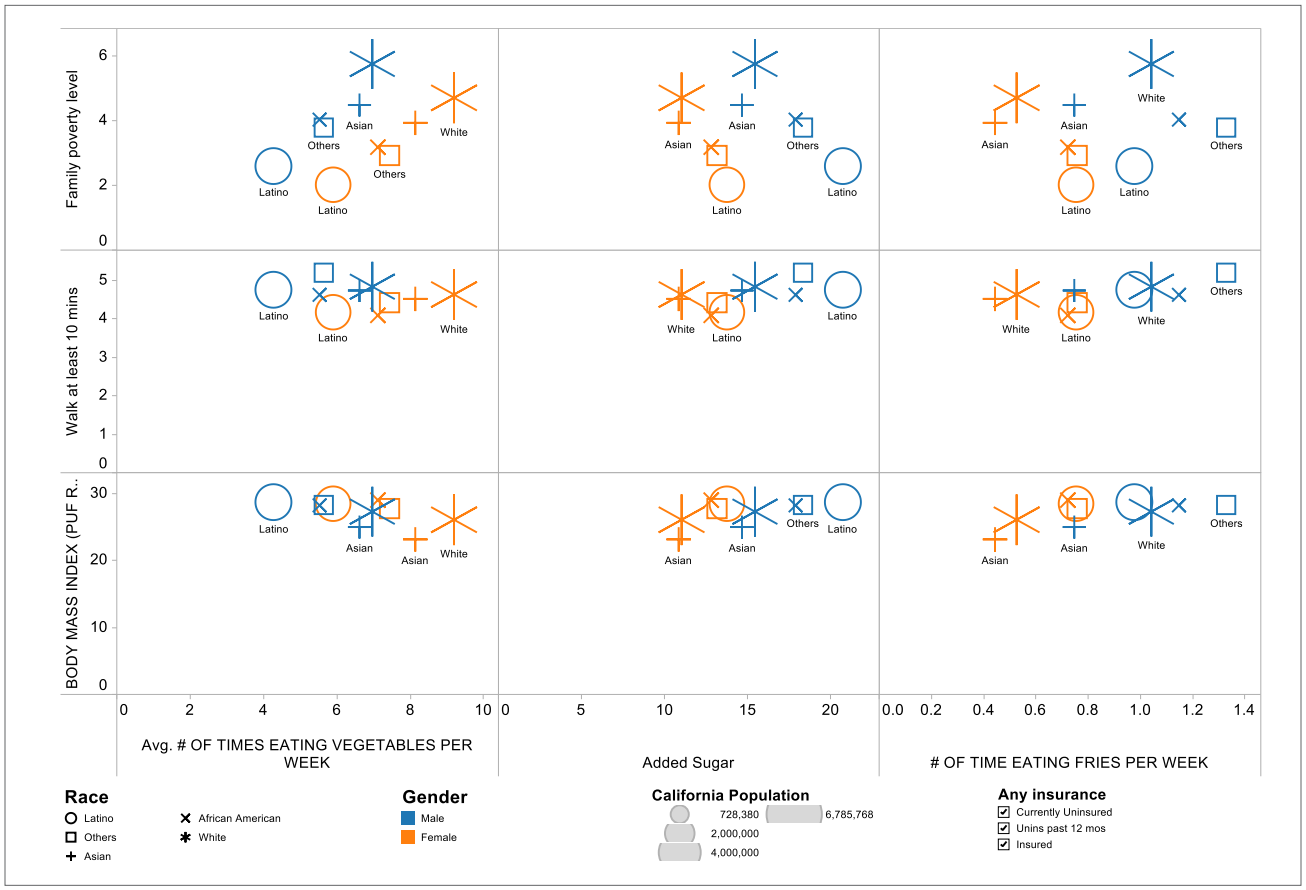


FIGURE 5: *Food preference, body mass index, physical activity, and poverty by race, California Health Interview Survey, 2009*

FIGURE 5 is an example of displaying multi-dimensional data. In this figure, we focused on patterns of food preference by race and ethnicity, given poverty level,[17] whether the respondent reported walking for at least ten minutes in the past seven days, and BMI. White women (orange stars) consume more vegetables than any other group (first column). We observe a linear relationship between income and vegetable consumption (upper left square), and the gender differences in consumption were consistent across all races. A similar but opposite relationship exists between sugar and income (middle column, top row). The gender pattern for fry consumption was less consistent than the pattern for vegetables and sugar (upper right corner). All groups shared similar average physical activity level (the second row) and BMI (third row).

DISCUSSION

In this paper, we demonstrated the relationships between obesity and a range of factors: gender, race/ethnicity, health behaviors, and diabetes. Our findings are consistent with much of the previous literature on obesity, such as gender and race differences in BMI, but also revealed some new relationships, such as the differences in the relationship between health behaviors and BMI by race/ethnicity. We were able to unveil these relationships easily and produce figures that can be quickly understood by a wide audience.

Visual analytics open up the possibility of rapid and intuitive analysis of complex, high-dimensional data to all users. In seeking important relationships within high dimensional datasets, visualizations exercises such as this one enable users to rapidly identify useful connections worthy of deeper analysis. In this exercise, we have visually identified various factors related to obesity that are worthy of further investigation. It would be very difficult to accomplish that using a traditional approach.

ACKNOWLEDGEMENTS

We appreciate the help and support of our employer, Blue Shield of California, and our leaders, Kimball Wilkins and Doug Biehn. We are not compensated for this article and declare no Conflict of Interest. All discussions are our own personal opinions.

BIOGRAPHY

Aaron Lai, CFA, is the Senior Manager of Marketing Analytics. He has extensive experience in consumer analytics with publications in both academic and professional journals. He has a M.Sc. in sociology and is studying for another M.Sc. in Evidence-based Healthcare both from the University of Oxford.

Thomas Ho is an experienced statistical modeler and is currently a Lead Marketing Analyst. His expertise is in scorecard development for consumer and commercial marketing and risk management.

Ryan Walker is a PhD candidate in Mathematics at the University of Kentucky.

Allison Sullivan is a survey analyst at Blue Shield of California. She has a Ph.D. in demography from the University of Pennsylvania. Her research focuses on the associations between socioeconomic status, social support, and mortality.

NOTES

1 "Obesity and Overweight for Professionals: Data and Statistics: Adult Obesity - DNPAO - CDC."

2 Flegal KM, Carroll MD, Kit BK, Ogden CL., "Prevalence of Obesity and Trends in the Distribution of Body Mass Index Among US Adults, 1999–2010." *JAMA*. 307, 5, (2012):491–497 (doi:10.1001/jama.2012.39).

3 Mehta N and Chang V., "Mortality Attributable to Obesity among Middle-Aged Adults in the United States." *Demography*. 46, 4 (2009): 851–872.

4 See note 2 above.

5 "About CHIS." http://askchis.com/about.html (accessed: September 2, 2012).

6 The public use data file used in this article is available from http://healthpolicy.ucla.edu/chis/.

7 Johnson, DW and Lee, W., *Explaining the Female Black-White Obesity Gap: A Decomposition Analysis of Proximal Causes.* Demography., 48 (2011):1429–1450.

8 Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of Overweight and Obesity in the United States, 1999–2004. *JAMA*., 295, 13 (2006), 1549–1555 (doi:10.1001/jama.295.13.1549).

9 Ibid.

10 Ibid.

11 See note 7, 8 above.

12 "The caveman's curse" The Economist. December 15, 2012, 5.

13 Mozzafarian, D, Hao, T, Rimm, EB, Willet WC, and Hu FB., "Changes in Diet and Lifestyle and Long-Term Weight Gain in Women and Men." NEJM 364:25 (June 23, 2011), 2392-2404.

14 See note 12 above.

15 The weight variable "RAKEDW0" in CHIS is the ample weight. The variables "RAKEDW1"- "RAKED80" are only required for computations of variances and standard errors.

16 For simplicity and to avoid sample size issues, we group underweight and normal individuals as normal BMI.

17 Poverty level indicates household income relative to the federal poverty guidelines; 1= below the poverty line, 2=100–199%, 3=200–299%, and 4=300% of the federal poverty line and above. The poverty guideline for a household of 4 was $22,050 in 2012.